5П

УДК 004.08

А.В. Палагин[1], *С.П. Риппа*[2], *А.А. Саченко*[3]
[1]Институт кибернетики им. В.М. Глушкова, Национальная академия наук Украины, г. Киев
[2]Национальный университет ГНС Украины, г. Ирпень
[3]Тернопольский национальный экономический университет, г. Тернополь
palagin_a@ukr.net, rippa_serg@ukr.net, as@tneu.edu.ua

Концептуализация и проблематика онтологий

Статья посвящена анализу и обобщению концептуальных основ применения онтологий в процессах создания информационных систем. Рассматриваются состав и архитектура онтологий как основы семантического Веба, в котором особое внимание уделяется понятию баз знаний как центрального компонента интеллектуальной информационной системы в Интернете. Приведены примеры выполненных и актуальных проектов по тематике онтологий, которые реализуются в академических лабораториях и вузах Украины, сформулированы также основные проблемы в сфере знание-ориентированных систем и направления их решения.

Теория искусственного интеллекта (ИИ) имеет многолетнюю историю и разветвленную архитектуру как наука. Следует отметить, что за последние 20 лет наиболее интенсивные изменения произошли в интеллектуальных науках тех направлений, которые соотносятся и соприкасаются с Интернетом. В названном контексте можно сформулировать основную **цель данной статьи** – проанализировать и обобщить новые концепции Интернета, проявившиеся в так называемом семантическом Вебе с позиций ИИ. Одним из способов достижения поставленной цели, на наш взгляд, может стать решение двух задач: первая – расширить понимание существующей теории ИИ, ее наиболее важных понятий и методологий в направлении семантического Веба, а вторая – осмыслить концепции и основные противоречия онтологических подходов как основной идеологии распределенных в Интернете знание-ориентированных систем.

Следует признать, что наука ИИ как совокупность многих его направлений развивается неравномерно и в последние годы для ряда традиционных разделов, если подойти критически, следует признать, что достижения довольно скромные.

При этом, как уже было отмечено выше, сегодня значительно активизируются исследования на стыке с Интернетом. Уверенно в научную терминологию вошли такие понятия, как «семантический Веб», «интеллектуальные сети», «распределенные информационные системы», «мультиагентные системы» и одновременно расширился теоретико-методологический аппарат знание-ориентированных наук и технологий. В данном контексте следует отметить направления исследований универсальных языков, XML-систем, таксономий и онтологий, которые в комплексе создали довольно стройную систему развития следующих поколений Интернета, включая идею семантического Веба, ставшего катализатором создания глобального информационного общества [1]. Реализацию семантического Веба можно считать практическим воплощением сетевой парадигмы искусственного интеллекта. Значение Интернет-сообщества в исследованиях фундаментальных проблем ИИ переоценить трудно, поскольку сложность и комплексность широкомасштабных научных работ требует такой же широкомасштабной координации, проводимой консорциумом W3C и начатой в свое время Тимом Бернерсом-Ли с официального обоснования семантического Веба [2]. В Украине развиваются старые признанные и создаются новые научные школы знание-ориентированных наук. Несмотря на минимальное финансирование и пассивность государства в плане поддержки исследовательских программ научные центры Киева, Харькова, Донецка за

последние 10-15 лет осуществили ряд серьезных теоретических и практических разработок в этой области. Особо интенсивное развитие получило направление формальных компьютерных онтологий.

Обобщение основных концепций онтологического подхода дает возможность осмыслить его роль и составные элементы в рамках общей теории ИИ. Упоминание роли онтологий относительно развития теории ИИ не случайно, поскольку именно онтологии по определению представляют собой методологию и инструмент описания сущностей и природы бытия, в том числе знаний и метазнаний. Понятие онтологии в информатике характеризует попытку всеобъемлющей и детальной формализации некоторой области знаний с помощью прозрачной логической концептуальной схемы. Применительно к ИИ онтология определяется как форма представления знаний о реальном мире или его фрагментов (что и зафиксировано в компьютерной энциклопедии [3]). Естественно, что строгого и единственно правильного определения онтологии мы не найдем, поскольку эта научная область в данное время интенсивно формируется силами многих ученых-информатиков со всего мира со своими субъективными точками зрения. В итоге можно собрать целое множество определений онтологии, которые частично пересекаются и во многом отличаются друг от друга, большей частью ориентируясь на отдельные стороны и специфические свойства интеллектуальных систем. При этом главным, на наш взгляд, в онтологическом подходе является то, что объединяет большинство разделов теории ИИ и этим общим является понятие базы знаний. Иными словами, онтологии позволяют формализовать и компактно представлять накопленные знания, одновременно определяя и объединяя терминологию различных предметных областей, строить единую научную картину мира как результат комплексных трансдисциплинарных исследований. Более того, в некоторых публикациях [4] онтологии, рассматриваемые в структуре семантического Веба, напрямую отождествляются с универсальными базами знаний. При таком подходе база знаний представляет собой описание понятийного аппарата определенных предметных областей, допускающее последовательные трансформации представлений знаний в формах, ориентированных на восприятие и использование мультиагентными интеллектуальными информационными системами.

Архитектура онтологии в обобщенном виде базируется на трех составных – универсальной языковой основе (XML), методологической модели описания информационных ресурсов (RDF) и, собственно, на языке описания онтологий (OWL). Представленная на рис. 1 модель базы знаний (БЗ) в виде онтологий демонстрирует их составные элементы и их место в Интернете. Графически в динамике показано, что семантический Веб является частью традиционного Интернета, его трансформацией через расширение сегмента онтологий. В свою очередь, онтологическая составная отображается на основе трехуровневой модели баз знаний в виде XML-, RDF- и OWL-компонентов, которые и являются приложениями в разных областях ИИ. Из рис. 1 видно, что допускаются также различные сочетания составных семантического и традиционного Веба, поскольку процесс трансформации и наполнения онтологий эволюционный и охватывает не только Веб, но и другие сегменты Интернета. В качестве такого примера может рассматриваться появление в группах новостей RSS-систем, которые также используют XML-базис.

Реализации онтологических проектов поддерживаются во многих исследовательских учреждениях и вузах Украины. В данном контексте можно назвать как научно-исследовательские, так и образовательно-учебные исследования, различие между которыми достаточно условно.

Так, в ИК НАНУ в течение ряда лет в рамках академической тематики развивается направление знание-ориентированных систем, в котором представлены принципы построения онтолого-управляемых информационных систем с реконфигурируемой архитектурой, методы онтологического анализа естественно-языковых текстов на предмет извлечения из них знаний, прикладные аспекты онтологий, в частности, для разработки электронных учебных курсов, метаонтологии и системы интеграции знаний в трансдисциплинарных областях и др. [5-7] [8].

Еще один проект по разработке Веб-онтологий как средств обмена данными и поддержки решений для улучшения экономического сотрудничества между Украиной и США финансировался NSF и выполнялся в 2004-2007 гг. НИИ интеллектуальных компьютерных систем (НИИ ИКС) Тернопольского Национального экономического университета совместно с New Jersey Institute of Technology, USA [9].

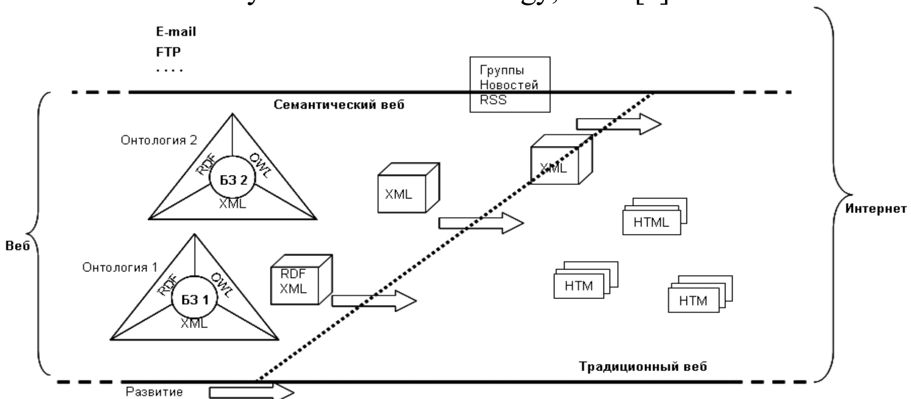


Рисунок 1 – Обобщенное представление архитектуры и составляющих семантического Веба

Следующих два актуальных проекта, о которых целесообразно упомянуть подробнее, осуществляются совместными усилиями Научно-исследовательского центра по проблемам налогообложения Национального университета ГНС Украины и НИИ ИКС: первый – создание системы электронного аудита налоговых деклараций [10] и второй – разработка методологии трансфера знаний в сфере Компьютинга (компьютерные науки, компьютерная инженерия, программная инженерия, информационные системы и технологии, в том числе и в экономике) для содействия мобильности студентов и внедрения дистанционного обучения в рамках международных образовательных стандартов [11].

В первом проекте усилия разработчиков сосредоточены на создании инструментальных средств поддержки в Интернете системы электронной налоговой отчетности и аудита на основе онтологий. Начиная с 2006 года, в системе налоговой службы Украины внедряется XML-ориентированная отчетность налоговых деклараций, используя которую в качестве первичных таксономий предметной области «Электронная налоговая отчетность и аудит», становится возможным описание информационных ресурсов деклараций и методики их аудита в Интернете (RDF-составляющая). База знаний первых двух уровней (XML- и RDF-описания) используется для формирования инструментальных средств формализации и поддержки онтологии «Электронного налогового аудита» в системе органов Государственной налоговой службы (ГНС), использование которых обеспечивает расчет и экспресс-анализ показателей налоговой нагрузки. Программная поддержка онтологии для оценки налоговой нагрузки осуществляется совокупностью мультиагентных модулей с применением Java-скриптинга для выполнения XSLT-преобразований. Особенностью реализуемого проекта является ориентация онтологии на проектирование всей системы в среде дистанционного обучения (СДО «Прометей»), которая, с одной стороны, содействует внедрению знание-базированной технологии реформирования деятельности органов ГНС, а с другой стороны, включает модули для плательщиков налогов, осуществляя процесс их обучения и адаптации к системе электронной налоговой отчетности [12].

Актуальность второго проекта подчеркивает тот факт, что одной из ключевых проблем для вхождения Украины в открытую американскую и европейскую образовательную среду является перевод украинских университетов на систему зачетных кредитов, способную обеспечить приемлемый способ измерения, и сопоставление результатов учебы в трансфертных процессах знаний для студентов и ученых при переходе и взаимодействии между разными национальными университетами и вузами. Кредитная система

ECTS (European Credit Transfer System) позволяет проводить структуризацию учебных планов университетов разных стран Европы с целью обеспечения их совместимости, однако, если сравнение дисциплин учебных планов с разными единицами измерения нагрузки можно достаточно легко осуществить с помощью предварительного перевода нагрузки в систему ECTS, то проведение качественной оценки является достаточно трудоемким процессом вследствие актуальности проблемы формализации данной предметной области. Анализ показывает, что одним из наиболее эффективных путей решения указанной проблемы является применение онтологий, активизация исследования которых стремительно нарастает в последние годы.

В процессе реализации указанного проекта предложен подход, основанный на сравнительном анализе содержания университетских учебных планов и программ по Компьютингу с использованием онтологии как инструмента их формализованного описания. При этом основное внимание уделяется не формальному перечислению кредитов из учебных планов с разными единицами нагрузки (как, например, ECTS), а определению количественной оценки сравнения дисциплин на основе анализа их содержания. Известно, что предметная область (область знаний) Компьютинга может быть разделена иерархически на подмножества разделов знаний (отдельных блоков знаний – дисциплин), которые, в свою очередь, дифференцируются на тематические модули (дисциплины), причем каждый модуль включает набор подразделов дисциплин (тем). Когда рассматриваем процесс разработки онтологии, в данном случае ключевыми стадиями являются концептуализация и формализация. На этапе концептуализации онтологии проводится содержательный анализ блоков знаний и определение основных понятий, их взаимосвязи и свойства (учебные планы, дисциплины, виды отчетности, учебной нагрузки, этапы учебы, блок дисциплин, структура и характеристики учебного плана). На этапе формализации онтологии выполняется формализация основных понятий и определяются способы интерпретации знаний. При этом для описания декларативных знаний представилось целесообразным использовать семантические сети и фреймы, а для представления процедурных знаний – процедурные и формально-логические модели с использованием методов концептуализации образовательных знаний и привлечением аппарата универсальных языков и таксономий. Таким образом, онтология отображает семантику области знаний, а сами семантические сети обеспечивают представления знаний в виде ориентированного графа, вершины которого содержат основные понятия из соответствующих дисциплин, а дуги иллюстрируют взаимосвязи между ними. Кроме того, предусмотрена разработка составной онтологии представления знаний по учебным планам, конечным этапом которой является описание методологии трансфера знаний по Компьютингу.

Описанный подход частично реализуется на сервере дистанционной учебы [12], который поддерживается совместно Национальным университетом Государственной налоговой службы Украины и Тернопольским национальным экономическим университетом. Ожидается, что результаты исследования будут иметь значение для сферы Компьютинга в системе высшего образования как Украины, так США и Европы, потому что они предоставляют формальный инструмент сопоставимости содержания учебных планов и программ дисциплин и, таким образом, обеспечивают возможность трансферта знаний в данной сфере вместо механического перезачета кредитов, как это осуществлялось ранее. Указанный подход позволяет интенсифицировать работы по согласованию образовательных стандартов Украины и США, а также стран Европы и, на основе семантического перезачета кредитов, придает более осмысленный характер мобильности студентов.

Исследовательский процесс онтологических подходов в теории знание-ориентированных наук, как любая исследовательская область, характеризуется множеством проблем и нерешенных задач. Одной из них является необходимость концептуализации метаонтологий или метазнаний в области ИИ. И действительно, за редким исключением, большинство проектов в сфере онтологий после декларирования принципов семантического Веба и построения его отдельных, условных фрагментов продолжаются традиционными средствами гипертекстовых систем и технологий программирования.

Другой нерешенной проблемой можно считать противоречие в сфере правового регулирования лицензионных вопросов создания и поддержки информационных и программных систем. Тут идет речь о конкурентной борьбе так называемых «открытых и закрытых систем». Общепризнано, что коммерческие лицензии на одноразовую покупку программного обеспечения является достаточно сильным сдерживающим фактором прогресса информационных технологий. Данную ситуацию иногда сравнивают с необходимостью для каждого нового велосипедиста повторно изобретать велосипед. Для научного сообщества очевидны преимущества открытых систем, которые совершенствуются и развиваются общими усилиями в среде глобального информационного общества. Сегодня уже понятно, причем не только интуитивно, но и по результатам развития знание-ориентированных наук, что различные методы регулирования лицензионных аспектов применения информационных систем, в том числе интеллектуальных, могут в значительной мере ускорять или замедлять развитие целых научных сфер.

Перечень методологических проблем и нерешенных задач в области онтологий можно продолжить. Например, отсутствие приемлемого теоретического обоснования инструментальной реализации онтологических систем, поскольку большинство проектов в этой сфере поддерживаются традиционными технологиями программирования и Веб. Похожие проблемы характеризуют теоретические аспекты создания интеллектуальных мультиагентных и Веб-сервисных систем, ориентированных на семантический Интернет. Таким образом, обобщение большинства названных проблемных вопросов дает возможность выделить ключевую сущность, объединяющую разнообразные онтологии в области ИИ. Эта сущность материализуется в понятии базы знаний, которая с расширением интеллектуальных технологий в Интернете, трансформацией его в семантический Веб, приобретает в такой новой Сети универсальные выразительные возможности, овеществляясь в онтологиях.

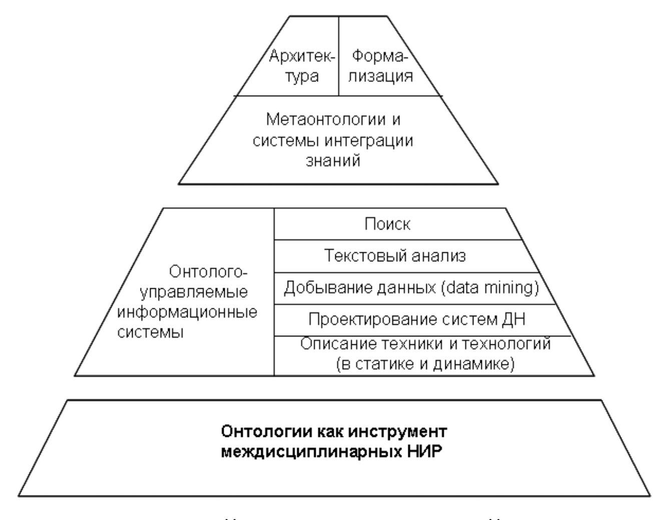


Рисунок 2 – Междисциплинарный характер онтологий в научных исследованиях

Подводя итог, можно представить онтологии на рис. 2 не просто в виде составных компонентов, а как универсальный инструмент междисциплинарных научных исследований в виде пирамиды наиболее важных обобщенных функций знание-базированных систем, которые были концептуально охарактеризованы на примере рассмотренных проектов или описаны в терминах проблем и нерешенных задач в сфере интеллектуальных систем. Средний (знания) и верхний уровни (метазнания) являются необходимыми и достаточными элементами целостной теории баз знаний, которая формируется как концепция онтологий и одновременно выступает действенным и практическим инструментом в современном Интернете, понимаемом и воспринимаемым уже как семантический Веб в широком междисциплинарном смысле.

Выводы

1. Показано, что исходной доминантой исследований семантического Веба как расширяющейся составляющей традиционного мультимедийного Интернет-контента стали теоретические положения онтологий и баз знаний. На этой основе обобщены концепции развития знание-ориентированных разделов наук ИИ и обосновано для них значение онтологий в трансдисциплинарном исследовательском контексте.

2. Предложена пирамидальная модель онтологического обеспечения исследований и разработок в сфере информационных систем и технологий, которая позволяет концептуализировать архитектуру семантического Веба на двух уровнях: на первом – знания в виде онтологоуправляемых информационных систем, на втором – метазнания в процессе интеграции онтологий и баз знаний. Представленная модель подкреплена перечнем выполненных и актуальных проектов в области семантического Веба, которые реализованы силами исследовательских лабораторий НАНУ и вузов Украины.

Литература

1. Закон України «Про основні засади розвитку інформаційного суспільства в Україні на 2007 – 2015 роки», № 537-V від 09.01.2007 р.
2. Тим Бернерс-Ли, Джеймс Лендлер, Ора Лассила. Семантический Веб. – Режим доступа: http://czolin.piscm.net/logic/semantic_web_rus.html.
3. Режим доступа: http://ru.wikipedia.org/
4. Ландэ Д. Семантический веб: от идеи – к технологии. – Режим доступа: http://poiskbook.kiev.ua/sw.html .
5. Палагин А.В. Архитектура онтологоуправляемых компьютерных систем // Кибернетика и системный анализ. – 2006. – № 2. – С. 111-124.
6. Палагин А.В. Модель категориального уровня языково-онтологической картины мира / Палагин А.В., Петренко Н.Г. // Математичні машини і системи. – 2006. – № 3. – С. 91-104.
7. Палагин А.В. Системная интеграция средств компьютерной техники / Палагин А.В., Яковлев Ю.С. – Винница, Универсум-Винница, 2005 – 680 с.
8. Palagin A.V. Development of procedures of recognition of objects with usage multisensor ontology controlled instrumental complex / Palagin A.V., Peretyatko V. // International Journal "Information Theories and Application". – 2006. – V. 13, № 4. – P. 307-315.
9. Pasichnyk R., Sachenko A. Semantic WEB-Search Developing by Problem-Oriented Ontology Means // Proceedings of the IEEE International Workshop IDAACS'2007. – 2007. – Dortmund, Germany. – P. 445-448.
10. Rippa S., Lendyuk T. Selection of Alternative Projects Using Data Mining. IEEE International Workshop, IDAACS' 2007, Dortmund, Germany, September 6-9. – 2007. – P. 550-554.
11. Anatoly Sachenko, Allen Tucker. Design of Modern Academic Program in Computing for Ukrainian Universities. Society for Information Technology and Teacher Education International Conference. – Vol. 2003. – Issue. 1. – 2003. – P. 2790-2791.
12. Веб-сайт системы дистанционного обучения Национального университета ГНС Украины, раздел курсов кафедры интеллектуальных систем принятия решений. – Режим доступа: http://pr.ufei.ukrsat.com.
13. Ріппа С.П. Бази знань в управлінні ІТ-проектами // Міжнародна наук.-практ. конф. «Управління проектами: стан і перспективи». – К.: НУК, 2005. – С. 143-145.

Стаття присвячена аналізу та узагальненню концептуальних засад застосування онтологій в процесах створення інформаційних систем. Розглядаються склад і архітектура онтологій як основи семантичного Вебу, в якому особлива увага приділяється поняттю баз знань як центрального компоненту інтелектуальної інформаційної системи в Інтернеті. Наведені приклади виконаних і актуальних проектів з тематики онтологій, які реалізуються в академічних лабораторіях і вузах України, сформульовані також основні проблеми у сфері знання-орієнтованих систем і напрями їх розв'язання.

Alexander Palagin, Sergey Rippa, Anatoly Sachenko

Concept and Problematic Issues in Ontology

This paper is devoted to the analysis and generalization of ontology concepts implementation in processes of information systems design. Ontology composition and architecture are considered here as basis of the semantic Web where a special attention is paid by an idea of knowledge base as a prime component of intelligent information system in Internet. The samples of completed and conducted projects in ontology area which are running in research labs of the academic institutions and universities in Ukraine. The main problems in knowledge-oriented systems and ways of their solving are formulated there.

Статья поступила в редакцию 26.08.2008.